

This Provisional PDF corresponds to the article as it appeared upon acceptance. Copyedited and fully formatted PDF and full text (HTML) versions will be made available soon.

The genetic heterogeneity and mutational burden of engineered melanomas in zebrafish models

Genome Biology 2013, **14**:R113 doi:10.1186/gb-2013-14-10-r113

Jennifer Yen (jy3@sanger.ac.uk)
Richard M White (whiter@mskcc.org)
David C Wedge (dw9@sanger.ac.uk)
Peter Van Loo (pvl@sanger.ac.uk)
Jeroen de Ridder (j.deridder@tudelft.nl)
Amy Capper (amy.capper@hgu.mrc.ac.uk)
Jennifer Richardson (jennifer.richardson@hgu.mrc.ac.uk)
David Jones (drj@sanger.ac.uk)
Keiran Raine (kr2@sanger.ac.uk)
Ian R Watson (irwatson@mdanderson.org)
Chang-Jiun Wu (cwu7@mdanderson.org)
Jiqiu Cheng (jiqiu.cheng@esat.kuleuven.be)
Iñigo Martincorena (im3@sanger.ac.uk)
Serena Nik-Zainal (snz@sanger.ac.uk)
Laura Mudie (lm5@sanger.ac.uk)
Yves Moreau (yves.moreau@esat.kuleuven.be)
John Marshall (jm18@sanger.ac.uk)
Manasa Ramakrishna (mr9@sanger.ac.uk)
Patrick Tarpey (pst@sanger.ac.uk)
Adam Shlien (as16@sanger.ac.uk)
Ian Whitmore (iw2@sanger.ac.uk)
Steve Gamble (sg8@sanger.ac.uk)
Calli Latimer (cdt@sanger.ac.uk)
Erin Langdon (langdone@mskcc.org)
Charles Kaufman (charles.kaufman@childrens.harvard.edu)
Mike Dovey (mike.dovey@gmail.com)
Alison Taylor (alison_taylor@dfci.harvard.edu)
Andy Menzies (am3@sanger.ac.uk)
Stuart McLaren (sm2@sanger.ac.uk)
Sarah O'Meara (som@sanger.ac.uk)
Adam Butler (apb@sanger.ac.uk)
Jon Teague (jwt@sanger.ac.uk)
James Lister (jalister@vcu.edu)
Lynda Chin (lchin@mdanderson.org)
Peter Campbell (pc8@sanger.ac.uk)
David J Adams (da1@sanger.ac.uk)
Leonard I Zon (zon@enders.tch.harvard.edu)
E Elizabeth Patton (elizabeth.patton@hgu.mrc.ac.uk)
Derek L Stemple (ds4@sanger.ac.uk)

© 2013 Yen *et al.*

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

P Andy Futreal (afutreal@mdanderson.org)

ISSN 1465-6906

Article type Research

Submission date 4 June 2013

Acceptance date 11 October 2013

Publication date 23 October 2013

Article URL <http://genomebiology.com/2013/14/10/R113>

This peer-reviewed article can be downloaded, printed and distributed freely for any purposes (see copyright notice below).

Articles in *Genome Biology* are listed in PubMed and archived at PubMed Central.

For information about publishing your research in *Genome Biology* go to

<http://genomebiology.com/authors/instructions/>

The genetic heterogeneity and mutational burden of engineered melanomas in zebrafish models

Jennifer Yen¹

Email: jy3@sanger.ac.uk

Richard M White²

Email: whiter@mskcc.org

David C Wedge¹

Email: dw9@sanger.ac.uk

Peter Van Loo^{1,3}

Email: pvl@sanger.ac.uk

Jeroen de Ridder⁴

Email: j.deridder@tudelft.nl

Amy Capper⁵

Email: amy.capper@hgu.mrc.ac.uk

Jennifer Richardson⁵

Email: jennifer.richardson@hgu.mrc.ac.uk

David Jones¹

Email: drj@sanger.ac.uk

Keiran Raine¹

Email: kr2@sanger.ac.uk

Ian R Watson⁶

Email: irwatson@mdanderson.org

Chang-Jiun Wu⁶

Email: cwu7@mdanderson.org

Jiqiu Cheng⁷

Email: jiqiu.cheng@esat.kuleuven.be

Iñigo Martincorena¹

Email: im3@sanger.ac.uk

Serena Nik-Zainal¹

Email: snz@sanger.ac.uk

Laura Mudie¹

Email: lm5@sanger.ac.uk

Yves Moreau⁷
Email: yves.moreau@esat.kuleuven.be

John Marshall¹
Email: jm18@sanger.ac.uk

Manasa Ramakrishna¹
Email: mr9@sanger.ac.uk

Patrick Tarpey¹
Email: pst@sanger.ac.uk

Adam Shlien¹
Email: as16@sanger.ac.uk

Ian Whitmore¹
Email: iw2@sanger.ac.uk

Steve Gamble¹
Email: sg8@sanger.ac.uk

Calli Latimer¹
Email: cdt@sanger.ac.uk

Erin Langdon⁸
Email: langdone@mskcc.org

Charles Kaufman⁸
Email: charles.kaufman@childrens.harvard.edu

Mike Dovey⁸
Email: mike.dovey@gmail.com

Alison Taylor⁸
Email: alison_taylor@dfci.harvard.edu

Andy Menzies¹
Email: am3@sanger.ac.uk

Stuart McLaren¹
Email: sm2@sanger.ac.uk

Sarah O'Meara¹
Email: som@sanger.ac.uk

Adam Butler¹
Email: apb@sanger.ac.uk

Jon Teague¹
Email: jwt@sanger.ac.uk

James Lister⁹
Email: jalister@vcu.edu

Lynda Chin¹⁰
Email: lchin@mdanderson.org

Peter Campbell¹
Email: pc8@sanger.ac.uk

David J Adams¹
Email: da1@sanger.ac.uk

Leonard I Zon⁸
Email: zon@enders.tch.harvard.edu

E Elizabeth Patton⁵
Email: elizabeth.patton@hgu.mrc.ac.uk

Derek L Stemple¹
Email: ds4@sanger.ac.uk

P Andy Futreal^{1,6,*}
Email: afutreal@mdanderson.org

¹ Wellcome Trust Sanger Institute, Cambridge CB10 1SA, UK

² Memorial Sloan Kettering Cancer Center and Weill-Cornell Medical College, New York, NY 10065, USA

³ Department of Human Genetics, VIB and University of Leuven, B-3000 Leuven, Belgium

⁴ Delft Bioinformatics Lab, Delft University of Technology, Delft 2628 CD, the Netherlands

⁵ MRC Institute of Genetics and Molecular Medicine, MRC Human Genetics Unit and Edinburgh Cancer Research Centre, Edinburgh EH4 2XU, UK

⁶ Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX 77054, USA

⁷ Department of Electrical Engineering, University of Leuven, B-3001 Leuven, Belgium

⁸ Dana Farber Cancer Institute and Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA

⁹ Department of Human & Molecular Genetics, Virginia Commonwealth University, Richmond, VA 23298-0033, USA

¹⁰ Institute for Applied Cancer Science, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

* Corresponding author. Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX 77054, USA

Abstract

Background

Melanoma is the most deadly form of skin cancer. Expression of oncogenic *BRAF* or *NRAS*, which are frequently mutated in human melanomas, promote the formation of nevi but are not sufficient for tumorigenesis. Even with germline mutated *p53*, these engineered melanomas present with variable onset and pathology, implicating additional somatic mutations in a multi-hit tumorigenic process.

Results

To decipher the genetics of these melanomas, we sequence the protein coding exons of 53 primary melanomas generated from several *BRAF*^{V600E} or *NRAS*^{Q61K} driven transgenic zebrafish lines. We find that engineered zebrafish melanomas show an overall low mutation burden, which has a strong, inverse association with the number of initiating germline drivers. Although tumors reveal distinct mutation spectrums, they show mostly C > T transitions without UV light exposure, and enrichment of mutations in melanogenesis, *p53* and MAPK signaling. Importantly, a recurrent amplification occurring with pre-configured drivers *BRAF*^{V600E} and *p53*^{-/-} suggests a novel path of *BRAF* cooperativity through the protein kinase A pathway.

Conclusion

This is the first analysis of a melanoma mutational landscape in the absence of UV light, where tumors manifest with remarkably low mutation burden and high heterogeneity. Genotype specific amplification of protein kinase A in cooperation with *BRAF* and *p53* mutation suggests the involvement of melanogenesis in these tumors. This work is important for defining the spectrum of events in *BRAF* or *NRAS* driven melanoma in the absence of UV light, and for informed exploitation of models such as transgenic zebrafish to better understand mechanisms leading to human melanoma formation.

Background

Melanoma is a form of skin cancer known for its therapeutic resistance, aggressiveness and late metastatic manifestation [1]. Activating mutations in *BRAF* (V600E) or *NRAS* (Q61K) are collectively found in approximately 60% of human melanomas and result in the constitutive signaling of the mitogen-activated protein kinase (MAPK) pathway [2,3]. Although studies have shown a clear dependence of tumour growth on MAPK signaling, most nevi with *BRAF*^{V600E} or *NRAS*^{Q61K} mutations remain benign for decades [4]. In zebrafish, expression of human *BRAF*^{V600E} (*BRAF*) or *NRAS*^{Q61K} (*NRAS*) in melanocytes results in the growth of pigmented, nevus-like lesions that also rarely progress to melanoma.

Invasive melanomas develop in these transgenic zebrafish only in combination with engineered loss of *p53* function [5,6], and yet manifest with variable onset and penetrance, strongly suggesting that these drivers are not sufficient for malignant melanoma formation and the requirement for additional unknown, somatic events.

Recent analyses of the genomes and exomes of human melanoma have resulted in the identification of new mutations that are likely to contribute to the disease formation or survival [7-11]. One confounding aspect of discriminating drivers in melanoma is the elevated background mutation burden due to UV mutagenesis, although new algorithms have been developed to refine this analysis [10]. We sought to build upon these studies through a focused analysis of a set of engineered melanomas, to determine the spectrum of mutations in the absence of UV light and to interrogate the role of *BRAF*, *NRAS* and *p53* in melanoma in transgenic zebrafish. Specifically, we used targeted exon enrichment and Illumina sequencing to generate exome and copy-number alteration data for 53 samples consisting of 38 *BRAF* driven and 15 *NRAS* driven primary zebrafish melanomas and cell lines with additional perturbations. A detailed examination of the spectrum of somatic point mutations, insertions, deletions and amplifications is presented. Our analysis reveals striking genetic heterogeneity, genotype specific mutation patterns and a potential novel path to *BRAF* driven tumorigenesis, providing insights into the events important for cooperation with *BRAF* and *NRAS* in the context of low mutation burden.

Results and discussion

Study set and sequencing overview

We collected matched zebrafish melanoma and normal tissue from 53 transgenic zebrafish harbouring tissue-specific oncogenic alleles of human *BRAF* and *NRAS* under a melanocyte-specific (*mitfa*) promoter [5,6] (Table 1, Additional file 1: Table S1, Figure 1). Specifically, 38 fish expressed oncogenic *BRAF*^{V600E} (*BRAF*) and 15 expressed oncogenic *NRAS*^{Q61K} (*NRAS*). The majority of samples (33 *BRAF* and 14 *NRAS* individuals) carried at least one germline, mutant *p53* allele (*p53*^{M214K} [12]). While *p53* itself has not traditionally been considered to be a major tumour suppressor in melanoma development, inactivation of *CDKN2A/p16* is associated with loss of *p53* activity [13]. Further, the high mutation load in *p53* and its pathway components in melanoma also underscores its importance [10]. Four *BRAF* fish harboured a germline temperature sensitive hypomorphic allele of *mitfa* (*mitfa*^{vc7}) [14,15]. Of *BRAF* individuals with aberrant *p53*, 38 had additional mutant germline alleles in *mitfa*^{-/-} (known as *nacre*^{-/-}) [16], *pten*^{hu1864 +/-} [17] or *mitfa*^{vc7} [14,15]. Transgenic individuals with *BRAF*;*p53*^{-/-};*mitfa*^{-/-} were manipulated with a miniCoopR shuttle vector system [18], consisting of somatic mosaic rescue of *MITF* expression in melanocytes along with *SETDB1* [18] and transcription factors *KROX20*, *FOXD3* or *OCT6*, the biology and oncogenicity of which are being investigated independently.

Table 1 Study set overview

Genotype	Samples
<i>mitf</i> : <i>BRAF</i> ^{V600E}	1
<i>mitf</i> : <i>BRAF</i> ^{V600E} ; <i>p53</i> ^{+/-}	2
<i>mitf</i> : <i>BRAF</i> ^{V600E} ; <i>p53</i> ^{+/-} ; <i>ptena</i> ^{hu1874 +/-}	1
<i>mitf</i> : <i>BRAF</i> ^{V600E} ; <i>mitf</i> ^{nc7 +/-}	4
<i>mitf</i> : <i>BRAF</i> ^{V600E} ; <i>mitf</i> ^{nc7 +/-} ; <i>p53</i> ^{+/-}	4
<i>mitf</i> : <i>BRAF</i> ^{V600E} ; <i>p53</i> ^{-/-} ; <i>mitf</i> ^{-/-} ; <i>mitf</i> : <i>MITF</i>	6
<i>mitf</i> : <i>BRAF</i> ^{V600E} ; <i>p53</i> ^{-/-} ; <i>mitf</i> ^{-/-} ; <i>mitf</i> : <i>MITF</i> ; <i>mitf</i> : <i>foxd3</i>	4
<i>mitf</i> : <i>BRAF</i> ^{V600E} ; <i>p53</i> ^{-/-} ; <i>mitf</i> ^{-/-} ; <i>mitf</i> : <i>MITF</i> ; <i>mitf</i> : <i>krox20</i>	12
<i>mitf</i> : <i>BRAF</i> ^{V600E} ; <i>p53</i> ^{-/-} ; <i>mitf</i> ^{-/-}	1
; <i>mitf</i> : <i>MITF</i> ; <i>mitf</i> : <i>krox20</i> / <i>foxd3</i> / <i>OCT6</i> *	
<i>mitf</i> : <i>BRAF</i> ^{V600E} ; <i>p53</i> ^{-/-} ; <i>mitf</i> ^{-/-} ; <i>mitf</i> : <i>MITF</i> ; <i>mitf</i> : <i>SETDB1</i>	1
<i>mitf</i> : <i>BRAF</i> ^{V600E} ; <i>p53</i> ^{-/-} ; <i>mitf</i> ^{-/-} ; <i>mitf</i> : <i>MITF</i> ; <i>mitf</i> : <i>EGFP</i>	1
<i>mitf</i> : <i>NRAS</i> ^{Q61K}	2
<i>mitf</i> : <i>NRAS</i> ^{Q61K} ; <i>p53</i> ^{-/-}	5
<i>mitf</i> : <i>NRAS</i> ^{Q61K} ; <i>p53</i> ^{+/-}	4
<i>mitf</i> : <i>NRAS</i> ^{Q61K} ; <i>p53</i> ^{+/-} ; <i>rps29</i> ^{+/-}	5
Total	53

Note that for asterisk (*) labeled tumours, the genes *krox20*, *foxd3* and *OCT6* were each expressed on separate plasmids.

Figure 1 Examples of zebrafish melanomas. *BRAF* (left panel) and *NRAS* (right panel) driven zebrafish melanomas in a *p53*^{-/-} background, with specimen example (top panel) and histology (bottom panel).

To analyze coding regions of the zebrafish genome, we performed targeted exome capture on tumour and normal DNA followed by 75 base paired-end Illumina (HiSeq) sequencing (European Nucleotide Archive accessions ERP003701, ERP003702). The bait set covered all protein coding genes, 3'UTRs and 5'UTRs of the Zv8 and later Zv9 genome for a combined coverage of 60 MB. A total of 2,309 Gb of sequencing was generated, averaging approximately 21.8 Gb per sample (Additional file 1: Table S2).

Because of the complexity and diversity of the zebrafish genome [19], we addressed the sensitivity and precision of applying the CaVEMan substitution calling algorithm [20] to zebrafish through two analyses: variant calling simulations and comparison to additional callers. In the first instance, we measured the performance of CaVEMan in simulated zebrafish tumour and normal genomes, which showed that the algorithm detected somatic substitutions with both high sensitivity and precision within these conditions (Additional file 2: Figure S1, Supplementary text in Additional file 3). We next employed CaVEMan for substitution calling on the zebrafish melanoma study set. Through manual inspection of each variant, we determined that a large proportion of these substitutions were false positives (57%, Additional file 2: Figure S2a), many due to germline variants that had been missed by the algorithm or calls made on suboptimal alignments (Additional file 2: Figure S2b). The low precision led us to manually examine all variants to ensure an accurate collection was used for downstream analysis.

In the second part of the analyses, we ascertained the sensitivity of our algorithm on the zebrafish melanoma dataset by comparing the CaVEMan calls for one sample (ZD8a) to those from SomaticSniper [21] and String Graph Assembler (SGA) [22]. Our results showed that SomaticSniper, and not SGA, provided a 10% increase of somatic variants to the CaVEMan algorithm (Additional file 2: Figure S2c-e). In spite of this marginal increase, we added a subset of non-overlapping Sniper variants to the CaVEMan calls, which we

experimentally validated through targeted enrichment and Illumina sequencing (Additional file 2: Figure S3). All calls from this analysis were then subject to a second, manual review.

Overview of substitutions and indels in engineered zebrafish melanomas

We confirmed a total of 403 point mutations and 13 insertions and deletions (indels), the latter of which were identified using Pindel [23] and processed using similar method to the substitutions (Figure 2a). Of the substitutions, 79 were synonymous, 168 resulted in amino acid changes, 16 were nonsense and 25 occurred at splice sites (Additional file 1: Table S3). Eighty-five substitutions were found in the 3'UTR and 26 in the 5'UTR, and one start codon was gained. The ratio of 2.3:1 non synonymous to synonymous events was similar to the averages previously reported in human melanoma [9,10]. The median number of coding mutations per sample was four, significantly fewer than the median of 171 in sun-exposed human melanomas and closer to the median of nine in mucosal and uveal melanomas, also originating from sun-shielded sites [9]. Over half of the total number of mutations in the study set was present in only eight samples (15%), six of which had two or fewer engineered “initiating drivers”. The highest number of substitutions were found in samples with one or two initiating drivers: ZD0038a (*BRAF*), ZD24a (*NRAS*), ZD23a (*NRAS*; *p53*^{+/-}) and ZD30a (*NRAS*; *p53*^{+/-}).

Figure 2 Overview of substitutions. A) The number of substitutions (dark blue columns) and indels (red columns) per sample, corresponding to their initiating germline mutations (bottom shaded). For *p53*, light blue indicates *p53*^{+/-} and dark blue *p53*^{+/-}. Asterisk (*) specifies *mitfa:MITF* expression in a *mitfa*^{+/-} background. **B-E)** Mutation spectrum of all and selected samples. For all samples (B) mutations are indicated on the transcribed (T) and untranscribed (UT) strand. **F)** Evidence of kataegis within 4,500 bp region in ZD8a, a *BRAF*; *p53* mutant sample. Somatic mutations are highlighted with coloured circles corresponding to the type of substitution.

Consistent with the low substitution burden, there were few recurrent mutations. Two substitutions were found in *ttna* and *ttnb*, the two largest protein-coding genes in the zebrafish genome. No recurrent substitutions were found in known melanoma genes or genes in the Cancer Gene Census [24]. Over 60% of genes mutated in this study were found to be mutated at least once in human melanoma [9-11,25-27], which was unsurprising given the extensive mutation load in the human disease. Substitutions with predicted coding changes in known census cancer genes included a nonsense mutation in *ikzf* and missense mutations in *nup214* and *pik3cd*, while a homozygous missense substitution in the anaphase promoting complex gene, *anapc1*, was identified in a *BRAF*, *p53*^{+/-} tumour (ZD8a).

UV-independent mutation spectra and mutational processes

Intriguingly, recent studies have shown that over half of the driver mutations in human melanomas do not bear the UV radiation associated signature [10]. To explore the nature of the non-UV events, we examined the mutation spectrum in the engineered zebrafish melanomas developed under conditions without detectable UV light, as determined using a standard laboratory photometer (International Light 1400). As with most human cancers, C > T substitutions (24.4%) constituted the prominent mutation class across all samples, including ZD8a and ZD24a (Figure 2c-d), which had substantial mutation burdens. Remarkably, ZD0038a, which had the highest substitution load (n = 47), consisted of mutations occurring exclusively at cytosine or guanine residues (Figure 2e), a mutation

signature that has not yet been described in human cancers. In this sample, all coding substitutions apart from one resulted in a predicted missense ($n = 21$) or a nonsense change ($n = 3$). To determine if this was the result of positive selection, we calculated the dN/dS ratio using a mutation-selection model. We found that the rates of missense and nonsense mutations for this sample were ~ 5.5 and ~ 9.8 times higher than expected by neutral evolution, respectively, a result unlikely in the absence of positive selection ($P_{dM/dS} = 0.030$ and $P_{dNS/dS} = 0.031$).

Similar to findings in non-sun exposed human melanomas [9], no significant bias of mutations was found in any class on any particular strand (Figure 2b). By comparison, a mutation strand-bias caused by transcription-coupled repair has been demonstrated in UV light induced melanomas, lung and breast cancers, all of which display the characteristic signatures of their respective UV, tobacco and DNA repair mutagens [9,28,29]. The absence of this signature in our samples suggests that these repair processes are not overt unless triggered by a selective, mutagenic pressure.

ZD8a, a *BRAF* and *p53* mutant, presented two microclusters of mutations. Twelve substitutions (40% of the total load) spanned exons within a 4,500 base pair (bp) interval of the *hoxd9a* and *hoxd10a* genes (Figure 2f), while a second cluster of five mutations was found within a 5 kb interval (Additional file 2: Figure S4). These microclusters were reminiscent of “kataegis”, hypermutated regions resulting from a single event [30]. A close examination of the reads revealed that the substitutions occurred in *cis*, had similar variant allele fractions and were mostly C > T transitions ($n = 12/26$) (Figure 2f). In human, patterns of kataegis have been proposed to be related to mutational processes of the AID/APOBEC family of enzymes, which modulates antibody diversification by deaminating cytidines to deoxyuridine within immunoglobulin genes [30,31]. Although APOBEC emerged only in primates, they are believed to have derived from the functionally conserved AID enzymes [32], which may provide the mechanistic origin of these clusters in zebrafish.

Insertions and deletions

Indels were sparse, with a total of 13 confirmed indels across the 53 samples (Additional file 1: Table S4). This is lower than the sample average of two to four indels in human melanoma [9]. Eight indels were single-base pair indels, and all 13 (< 5 bp) were flanked by tandem repeat sequences on either side, evidence of a lapse in post-replicative mismatch repair found commonly in breast cancer genomes [30]. Ten indels were out of frame and likely to cause loss of gene function. Four indels (36%) were found in a sample mutant only in *BRAF* (ZD0038). Interestingly, a single nucleotide deletion resulting in a frameshift mutation was found in *pik3ip1* (V170fs*), which in human directly binds to the p110 catalytic subunit of PIK3 and negatively modulates its activity [33]. Its occurrence in a *BRAF*, *mitfa*^{-/-}, *p53*^{-/-} mutant sample is consistent with a role for PI3K cooperation with MAPK deregulation in human melanoma [34].

Overview of copynumber changes

In total, 991 amplification segments (copy number ≥ 5 for samples with ploidy < 2.7 , and copy number ≥ 8 for samples with ploidy ≥ 2.7) and 436 segments of homozygous deletions (copy number = 0) were identified by ASCAT [35]. There was marked variation in the number of copy number changes among samples in the study set, with a cumulative 5 Gb of losses or gains manifesting in over half of tumours analyzed. For samples represented by

both aCGH and ASCAT data, the frequency recurrence profiles of copy number changes from ASCAT generally agreed with those from aCGH performed on the same DNA stock (Additional file 2: Figure S5).

While the majority of samples (85%) harboured at least one amplification, only 30% of the samples showed any homozygous deletions (Additional file 1: Table S5). It is therefore worth noting that *BRAF* driven tumours mutant in *mitfa*^{vc7} had significantly more homozygous deletions than expected by chance ($P = 0.01$ by Chi-Square test, Figure 3b). *NRAS* subtypes, by contrast, did not reveal apparent commonalities (Figure 3a). Clustering of ASCAT and aCGH segments from all samples also did not reveal any regions of subgroup affiliation apart from the strong amplified signal on chromosomes 18 and 19 (Additional file 2: Figure S6), the latter of which is believed to be associated with the *BRAF*^{V600E} transgene integration as a concatemer on chromosome 19.

Figure 3 Identification of a frequently amplified locus on chromosome 3. Frequency profiles of tumours mutant in **A)** *NRAS*; *p53*^{-/-}; *X*, **B)** *BRAF*; *mitfa*^{vc7}; *X* tumours and **C)** *BRAF*; *p53*; *mitfa*^{-/-}; *mitfa*:*MITF*; *mitfa*:*X*, where *X* can include additional drivers as mentioned in the text. **D)** Amplification segments supporting a peak on chromosome 3 in tumours of *BRAF*; *p53*; *mitfa*^{-/-}; *mitfa*:*MITF*; *mitfa*:*X* background derived from exome sequencing (maroon segments) and aCGH (green dotted segments). Samples mutated are represented by inverted, colour-coded triangles above the corresponding gene indicated by the thick black bar. **E)** Frequently amplified genes in the entire dataset. **F)** Number of copies (Y-axis) of the genes (X-axis) in the region of amplified locus. Each line represents a tumour that is colour-coded according to either *BRAF*; *p53*; *mitfa*^{-/-}; *mitfa*:*MITF*; *mitfa*:*X* (yellow) or other (blue) background status. The most frequently amplified genes are highlighted in yellow for in Figures D-F.

Identification of a recurrently amplified region in a subset of zebrafish melanomas

A particularly striking finding was the recurrence of a 175 kb amplicon on chromosome 3 (50.0 to 51.2 Mb) in 10 tumours belonging to the *BRAF*, *p53*^{-/-}, *mitfa*^{-/-} background with *MITF* rescue (Figure 3c). Although this subgroup is the largest of our dataset (47%, 25/53), the clustering of the recurrent amplicon in this subgroup was unlikely to have occurred by chance ($P = 0.000256$ by Chi-Square Test). Amplified segments were supported by both ASCAT and aCGH (Figure 3d). The most frequently amplified genes were *prkacaa* and *samd1* (1 of 2) presenting in 10/53 samples, followed by *asf1ba* (n = 9), *wu:fj41e11* (n = 9) and *tecra* (n = 8) (Figure 3e). While amplifications were found across all 10 samples for *samd1* and *prkacaa*, they presented in five or fewer samples for flanking genes *RNF222* and *gcgr* (Figure 3d).

A simulation was performed to determine the likelihood of the events occurring in these genes, at this frequency, by chance. For each sample, the number and lengths of the amplified segments were randomly introduced across the target exome regions one million times, producing a *P-value* that was adjusted by Bonferroni correction (n = 6,677). We did not factor causes of amplification other than those by chance, such as nearby fragile sites, for which little information is available for zebrafish.

Based on our simulations, all genes recurrently amplified in six or more samples were likely to be significant, including *prkacaa*, *samd1*, *asf1ba*, *wu:fj41e11* and *tecra* (n = 13, Additional

file 1: Table S6). These genes did not show evidence of amplification or overexpression in human cancer datasets (CCLE, Oncomine, COSMIC) or large, comprehensive melanoma studies [9,10]. Genes recurrently amplified in fewer samples also showed significant enrichment ($P = < 0.05$, Additional file 1: Table S6). Among these, interestingly, was *tert* ($P = 0.0$, $n = 4$ samples), which encodes the reverse transcriptase subunit of telomerase responsible for maintaining the ends of chromosomes. *Tert* was the only known cancer gene recurrently mutated in our study set. In human melanoma, *TERT* is amplified [10,36] and harbours promoter mutations in as many as 90% of melanoma cases [7,8].

Identification of few recurrent homozygous deletions

A total of 366 deletion events were identified, affecting the same genes in at most three samples in the study set (Figure 4a). By performing the above simulations, we determined the majority (28/30) of genes deleted in three samples were unexpected by chance (Additional file 1: Table S7). The genes *nitr1i*, *nitr3a*, *nitr7b* and *nitr7a* were in a locus deleted in three samples belonging to both *BRAF* and *NRAS* mutant lines (Figure 4b). The *nitr* genes are members of a highly diversified, multigene family of novel immune type receptor found in teleosts. *Nitr* genes do not rearrange like immune receptors but show structural similarities to both the mammalian T-cell or Ig-like receptors [37,38]. Loss of these genes could be relevant to one facet of progression which is to avoid immune surveillance, consistent with a critical role of immune regulation in human melanoma [39]. Other recurrently deleted genes include *sema6d*, *plcd3a*, *mrps5*, *cyp2y3* and *xirp* (Figure 4c-h), none of which had been previously implicated in human cancer. Further investigation would provide insights into the contribution of these genes to tumourigenesis in zebrafish.

Figure 4 Overview of homozygous deletions. **A)** Frequency of homozygously deleted genes across samples. **B)** Recurrently deleted loci occurring in at least three samples that are driven by *BRAF* (dark blue) or *NRAS* (light blue), and the corresponding genes in these regions (right hand side). **C)** Examples of deleted segments (dark blue bars) and the genes in these regions (bottom labelled), represented by their exon structures (dark blue lines).

Relationships between age, drivers, and mutation burden

We exploited the model system to explore the footprints of mutagenesis assuming a uniform basal mutational clock. Where data for the age of fish (at tumour collection) was available, we found a significant correlation between age and the number of substitutions using the Pearson's Correlation test and a Generalized Linear Model ($R = 0.37$, $P = 0.02$, $GLM P = 0.0035$). Positive correlations have similarly been found in human melanoma [9]. Age of onset and number of drivers were themselves strongly negatively correlated. If each germline driver was counted as one event in a requisite multistep process, we could attempt to delineate a relationship between these initiating events and extent of the mutations. For transgenic zebrafish of the genotype *BRAF*^{V600E}; *p53*^{-/-}; *mitfa*^{-/-}; *mitfa:MITF*, a value of four drivers were assigned, due to the yet unknown contributions of the additional genes (*KROX20*, *FOXD3*, *OCT6*) to melanoma. Interestingly, our data showed a significant, inverse association between the number of drivers and the substitution events ($R = -0.45$, $P = 0.00075$, $GLM P = 0.00031$), indicating that a greater number of drivers require fewer additional events to generate the melanoma lesions. To determine if this correlation extended to copy number events, we considered each amplified or deleted segment as an event in the tumour. Our data showed that if we considered drivers and age together, this was also a

significant predictor of the total number of copy number events (*GLM* $P = 0.00011$, Additional file 1: Table S8).

Functional categorization of frequently mutated genes

Similar to human cancers, the engineered melanomas overall displayed high heterogeneity, where the majority of genes mutated in only one sample (68%) (Additional file 2: Figure S7a). Taking into account all the different modalities of mutation and their frequencies of occurrence, a *P-value* was calculated (using a binomial test) for each mutated gene (Additional file 1: Table S9). Due to the high frequency of recurrence, *prkacaa* and *samd1* presented with the highest significance ($P = 2.31 \times 10^{-8}$). Following this, we explored the potential functional themes underlying these aberrations through a KEGG pathway analysis, which revealed that the enrichment for most pathways declines when the minimum threshold for number of mutated samples is raised (Additional file 2: Figure S7b). From this we infer that although many genes are not frequently mutated and significant by themselves (i.e. mutated only once), the pathways in which they reside are significantly mutated. Among the enriched pathways in this study (Additional file 1: Table S10), two include biological processes that have been directly linked to the hallmarks of cancer (apoptosis and VEGF signaling, for angiogenesis), while deregulation of two other pathways, p53 signaling and melanogenesis (Additional file 2: Figure S7c), have previously been implicated in melanoma [40]. Also showing significant enrichment was the MAPK signaling and cell cycle pathway, thus further supporting functions important in human melanoma development.

Conclusions

We have provided a comprehensive overview of the genetic events in engineered zebrafish models harbouring known driver alleles. Several new insights into the mutagenic processes in non UV-mediated, engineered melanomas, and the biology of *BRAF* and *NRAS*-driven malignancies, can be drawn from these 53 exomes.

Our results show that in the absence of direct UV light, engineered melanomas develop similar mutational signatures to most human cancers, dominated by the evolutionary conserved spontaneous deamination of cytosine to thymidine [9]. We also found rare cases exhibiting remarkably distinct mutation spectra, including indications of kataegis and a novel, unclassified mutational signature.

Importantly, our results demonstrate that tumours driven primarily by a greater number of known cancer genes typically manifest with fewer mutations, suggesting that such models can be used to bound and estimate the number of events in human cancers. Mouse models of acute myeloid leukemia and mammary tumours similarly displayed fewer mutations and structural rearrangements, respectively, than their human counterparts [41,42]. It could be speculated that predisposed human individuals would support the same conclusion. With nearly half of the samples presenting no substitutions or indels, however, these two classes of mutations are unlikely to be the only route to the additional mutations needed for full melanoma development, with potentially other factors such as chromatin modifications at play.

The highly recurrent amplicon in transgenic lines with *BRAF;p53;mitfa^{-/-};mitfa:MITF* encompassing the genes, *prkacaa*, *samd1*, *tecra*, *wu:fj41e11* and *asf1b*, indicates a strong

selection for genes in this interval in mutant *BRAF*, *p53* and *mitfa* lines with *MITF* rescue. Although the amplicon was exclusive to the *BRAF;p53;mitfa^{-/-};mitfa:MITF* transgenic models, it is unclear whether it would also present in the *BRAF* or *BRAF;p53* mutant models given a larger sample cohort. Since *MITF* serves as a functional rescue in this transgenic line, the genetics of this subset may be comparable to human tumours that show dependency on *MITF* for growth, either through *MITF* amplification or over expression. That none of the genes in this amplicon have been reported so far as mutated in human melanoma could therefore be due to its specific occurrence with *BRAF*, *p53* and amplified *MITF*, found in less than 5% of *BRAF* mutant metastatic melanomas and a rare combination (Additional file 2: Figure S8).

In this amplicon, amplification of *prkaca*, which encodes one of two principal catalytic (C) subunits of protein kinase A (*pka*), is intriguing for several reasons. Human *PRKACA* is the principle catalytic subunit of protein kinase A (PKA) [43]. Although not previously associated with melanoma, the cAMP-PKA pathway is a major signal transduction pathway for melanin production, melanocyte proliferation and differentiation (reviewed in [44]) and has been implicated in pituitary tumorigenesis [45,46]. Mutations in *PRKARIA*, a PKA-regulatory subunit, cause an inherited syndrome called the Carney complex, characterized by pigmented skin lesions, schwannomas, recurrent mucocutaneous myxomas and endocrine neoplasms [47,48]. Indeed, cAMP-dependent PKA activation has been shown to result in the upregulation of the *mitfa* promoter, tyrosinase expression and melanin synthesis, affecting skin pigmentation and melanogenesis [49]. Of interest, recent data has interestingly demonstrated a link between pigment production and UV-independent melanomagenesis, where harmful accumulation of pheomelanin intermediates or by-products during pigment synthesis can promote tumour formation [40]. Thus, a potential consequence of *PRKACA* amplification may be disruption of PKA signaling and pigment production, pointing to its possible contribution to aberrant pigment production in UV-independent carcinogenesis.

An important observation of this study is that, apart from the amplicon, the *BRAF* and *NRAS* driven melanomas display striking genetic heterogeneity similar to human cancers and mouse cancer models [41,42]. One interpretation of this finding is that tumorigenic processes are achieved through the contribution of many different mutated genes, in line with previous findings in mice cooperativity screens [50] and low frequency drivers unveiled from emerging studies of human melanoma [9-11]. The enrichment of mutations in pathways known to be important for melanoma development, such as *MAPK* and *p53* signaling, in the presence of germline mutations affecting *BRAF*, *NRAS* and *p53*, also suggest that further modulation of these pathway signaling is required for full manifestation of the tumours.

To the best of our knowledge, the spectrum of somatic coding mutations in an engineered model of melanoma has not yet been described. The integrated analysis we report here thus provides a glimpse into the genetic paths to *BRAF* and *NRAS*-driven tumorigenesis, providing a framework for genomic characterization, and a standard for evaluating and prosecuting detailed biological questions in engineered animal models of cancer.

Materials and methods

Simulation of zebrafish cancer genomes

Individual zebrafish genomes were created with a SNP density of 0, 0.001, 0.01, 0.1, and 0.5 SNPs/base by randomly generating substitutions across the genome using an in-house simulation script. Using each individual genome, referred to as the “normal”, we created a second genome containing an additional 2,000 substitutions for the “tumour”. For each normal and tumour genome, we simulated 75 bp reads in FASTQ format using wgsim [51], specifying null for the base mutation rate, error rate and indel mutation rate. To simulate normal contamination, we combined normal and tumour FASTQ files for each individual according to the following proportions to obtain an average sequencing coverage of 80X (Table 2).

Table 2 Metrics for simulating normal contamination in tumour and normal genome FASTA files

SNP density	Type	Millions of reads			
		Normal	Tumour content		
			30%	60%	100%
0	Normal	8	5.6	3.2	0
	Tumour	0	2.4	4.8	8
0.001	Normal	8	5.6	3.2	0
	Tumour	0	2.4	4.8	8
0.01	Normal	8	5.6	3.2	0
	Tumour	0	2.4	4.8	8
0.1	Normal	8	5.6	3.2	0
	Tumour	0	2.4	4.8	8
0.5	Normal	8	5.6	3.2	0
	Tumour	0	2.4	4.8	8

The simulated tumour and normal pairs were subsequently processed through the Cancer Genome Project Sequencing Pipeline.

Sample collection

Zebrafish tumour and normal tissue samples were obtained from Amy Capper and Jennifer Richardson (Elizabeth Patton's lab, University of Edinburgh, Edinburgh, UK), and from Richard White and Charles Kaufman (Len Zon's lab, Boston Children's Hospital, Boston). All samples were obtained in accordance with the UK Home Office regulations, UK Animals (Scientific Procedures) Act 1986, and reviewed by the Wellcome Trust Sanger Institute Ethical Review Committee. Samples from Elizabeth Patton's lab were subject to histopathological review by a clinical pathologist (Marie Mathers, Edinburgh Western General Hospital). We were unable to perform histopathology on samples from Len Zon's lab. Normal tissue included sections from the fin, head, or gut. Zebrafish melanoma and normal DNA were extracted from fresh frozen tissues using the Qiagen Blood and Tissue DNAeasy Kit (Cat # 69504). Melanomas were derived from transgenic zebrafish expressing either the *BRAF*^{V600E} or *NRAS*^{Q61K} human oncogene as previously described [5,6,18].

Exome bait set

Exon sequences for bait set design were initially downloaded from BioMart [52] to encompass all protein coding genes, 3'UTR and 5'UTR regions from Ensembl 58 of the Zv8 genome. The bait set was subsequently adjusted to encompass additional genes from Ensembl 61 and new releases of the Zv9 genome (Zebrafish Agilent All Exon SureSelect). A total of 2,309 Gb of sequencing was generated, averaging approximately 21.8 Gb per sample, of which 79.6% of reads mapped and 55% of which mapped to target coding regions (i.e. “on target”) (Additional file 1: Table S2). By comparison to the human exome [20], the performance of the zebrafish exome was slightly lower (in human, 89% of reads map, averaging 62% on target coverage), requiring a greater total sequencing depth to acquire the desired baseline coverage of 20X.

DNA and library preparation, capture and sequencing

DNA libraries were prepared using the Illumina Paired End Sample Prep Kit according to the manufacturer's protocol. For targeted enrichment, in the first iteration, we designed a custom bait set to target the zebrafish exome for solution capture to include all the exons of all protein coding genes in the Zv8 Ensembl 58 gene build. Subsequently, an additional 2,059 genes were added to include improved annotations in the Zv9 assembly and Ensembl 59 gene build. Targeted enrichment was performed as described [53] following the manufacturer's instructions.

Sequencing with 75 base paired-end reads of targeted-enrichment libraries was performed on the Illumina GAIIx and the HiSeq 2000 sequencers. Reads were mapped to the zebrafish reference (Zv9 Ensembl 61) using the Burrows-Wheeler Algorithm (BWA version 0.5.9) [54] under default parameters and excluding library PCR duplicates.

Identification of substitutions variants

CaVEMan

CaVEMan (cancer variants through expectation maximization), an in-house algorithm, was employed to call single nucleotide substitutions in our dataset. Post-processing filters developed for human variant calling and additional filters were applied to the set of initial CaVEMan mutation calls to improve the specificity of the output.

SomaticSniper

Tumour and normal BAM files were processed by SomaticSniper [21] with a specification for read and base quality of at least 40. Raw variants were post-processed using scripts obtained through Github [55], modified to include a variant allele frequency of no more than 3% in the normal sample and less than 10% of the tumour, and without germline SNPs or indels within 5 bp of any of the normal zebrafish exomes. Variants were annotated using the Ensembl variant effect predictor (Ensembl 64 gene build) specifying only coding variants as output.

SGA

SGA analysis was run by Jared Simpson using a modified algorithm [22].

Identification of insertions and deletions

Insertions and deletions were called using a modified version of Pindel [23] as previously described [28]. To improve the identification of high confidence variants, we specified a requirement for a minimum depth of 15 reads in both tumour and normal samples. For small indels, at least four reads supporting the variant seen by Pindel and at least one by BWA were required. Larger indels were defined in non-repeat regions where the mutation was seen once on either strand by Pindel. All indels were manually reviewed for confirmation.

Variant validation

Capillary and 454 resequencing

Validation of substitutions and indel variants was initially attempted through capillary or 454 Roche resequencing of amplified PCR products spanning the mutation in the tumour and the normal DNA, which had been subject to whole-genome amplification from the original stock using GenomiPhi (illustra GenomiPhi HY DNA Amplification Kit, Cat# 25-6600-20), according to the manufacturer's instructions. Nested PCR improved PCR yield over a single round of amplification, but both capillary and 454 Roche approaches proved problematic in PCR-amplified zebrafish DNA.

Targeted capture and illumina sequencing

To circumvent problems with PCR-based validation, we designed a custom bait set targeting the mutant alleles for enrichment followed by Illumina sequencing. We streamlined the validation study set by qualitatively reviewing each variant and keeping only CaVEMAN calls that did not show germline mutations and were supported by high quality mapping reads and alignment. An additional 1,700 non-overlapping, Sniper variants (60% of the total non-overlapping Sniper calls) were selected at random to include in the validation set, comprising a bait set of 1.4 MB with minimal tiling probes flanking 60 bp on either side of each variant. DNA libraries were made as described above and pooled into eight samples per group with barcode identifiers. Targeted capture was performed with each pool according to manufacturer's instructions followed by 100 bp paired-end sequencing on the Illumina HiSeq 2000 and default BWA alignment. Mutant variants were confirmed on Samtools Pileup files using a separate, in-house validation script based on tumour and normal allele depth and quality. All confirmed variants were subjected to an additional, manual review.

Identification of copy number variants

Copy number variation was determined primarily through ASCAT [35]. Only segments under 10 MB in length were considered. Genes falling in these segment regions were annotated using the Ensembl variant effect predictor (Ensembl 64). Segment data was analyzed using R, Nexus Copy Number Software 6.1 (Biodiscovery, [56]), visualized using IGV [57,58] and plotted using Progenetix [59].

Array CGH

Array CGH (aCGH) was performed on a subset of 24 zebrafish melanoma normal and tumour samples using a Nimblegen Custom Design 12×135 K CGH Array (Roche Nimblegen Technologies, Cat #05223881001) containing 135,000 probes covering the length of the zebrafish Zv9 genome. In brief, tumour and normal DNA were labeled, competitively hybridized to the array for 48 hours, washed and scanned using a 5 micron scanner (Molecular Devices). Signal intensities were extracted using the DEVA v1.2.1 Software (Nimblegen, [60]). Overall data quality was evaluated as recommended in the DEVA Software User's Guide [61]. Segmentation was performed using the R Copynumber Package [62] and visualized using the Nexus Copy Number Software (6.1) (Biodiscovery), IGV [57,58] and Progenetix [59].

Statistical analyses

Codon selection

We used the method described in [63] to evaluate whether amino acid changes in ZD0038a occurred at a higher frequency than expected in the absence of positive selection. Briefly, we used 12 parameters to describe the different rates of the 12 possible single nucleotide substitutions, and two parameters (analogous to dN/dS) to describe selection at missense and nonsense mutations. This allowed us to quantify the strength of the selection without the confounding effect of sequence composition and different rates of each substitution type. Maximum-likelihood was used to estimate these parameters and Likelihood Ratio Tests were used to test deviations from neutrality (dN/dS = 1). Analogous results to those presented in the main text were obtained using the traditional codon model approach used in phylogenetic analyses (implemented in [64]) as well as accounting for CpG context-dependent effects.

Estimation of the number of mutated copies

Allele-specific copy number estimates for point mutations were obtained by integrating copy number and sequencing data as described in [35].

Simulations of amplifications and homozygous deletions

Simulations were performed as described in [65] (in preparation).

Evaluation of driver and age correlation

Mutation burden and driver correlation was performed as previously described [20].

Pathway analysis

Mutation significance analysis

We combined mutation data from substitutions, insertions and deletions, and copy number changes (amplifications and homozygous deletions) to assess the likelihood of a gene being mutated in more samples than expected by chance. As each mutation type can occur at a different frequency (where amplifications are more frequent than deletions), each mutation

type was considered separately. Thus, to calculate a combined *P-value* for each gene *j*, we used the following Equation 1:

$$p_j = \prod_{i \in \{1,2,3,4\}} P(X \geq x_{ij} | q_i, n_i) = \prod_{i \in \{1,2,3,4\}} \left(1 - \sum_{k=0}^{x_{ij}-1} \binom{n_i}{k} (q_i)^k (1-q_i)^{n_i-k} \right), \quad (1)$$

where x_{ij} is the number of samples carrying a mutation in gene *j* in sample group *i* and n_i the number of samples in sample group *i*. Moreover, q_i was calculated as follows:

$$q_i = \frac{1}{n_i} \sum_{k=1}^{n_i} 1 - \left(1 - \frac{1}{N} \right)^{m_k},$$

with *N* number of genes in the genome.

Entrez gene mapping

For compatibility with the KEGG database, we mapped *Danio rerio* Ensembl IDs onto Entrez IDs using NCBI [66], which includes a cross-reference of Entrez to Ensembl. Target genes that could not be matched in this fashion were matched using gene symbol and synonyms.

Pathway analysis

We used knowledge from the KEGG database to construct a large protein interaction network. To gauge whether a pathway contains more frequently mutated genes than expected by chance, a KEGG pathway enrichment was performed for all 215 pathways in the *Danio rerio* specific KEGG database.

All genes with a combined *P-value* < 0.05 (as calculated according to the Mutation Significance Analysis) were selected for the pathway analysis. This cutoff selected for genes with at least two amplifications, and given rarer mutation types, genes with at least one mutation other than an amplification.

We called a gene frequently mutated if it carried at least *N* mutations, where *N* can be between one and 10 (Additional file 1: Table S9). Genes with mutation counts of three or more were visualized in the context of their KEGG pathway interactions using Cytoscape [67].

Competing interests

The authors declare no competing interests.

Authors' contributions

JY designed the study, performed research, analyzed data and wrote the paper. RW, EEP, DS and PAF participated in the study design. DCW, PVL, JDR, IM, IW, CJW analyzed data and performed research. RW, AC, JR, CKK, EL, AMT, LC, LZ, JL, DJA, EEP contributed novel reagents. MD, JC and YM contributed graphic support. DRJ, JM, PT, KR, AB, JT, SM

contributed to the data processing. SZ, MR, AS, PT contributed to the data analysis. LM, IW, SG, CL, SO contributed to the labwork. RW, DCW, PVL, JDR, IM, CK, DJA, LIZ, EEP, DS and PAF critically read the manuscript, which was approved by all the authors.

Acknowledgements

We would like to thank Jared Simpson for running the SGA analysis and Bettina Fischer for help with the aCGH. This work was supported by the Wellcome Trust. PVL is supported by the Flanders Research Foundation (FWO); AS by a HL Holmes Award from the National Research Council and DJA by Cancer Research UK; CKK by NIH grant K08AR61071.

References

1. Flaherty KT, Hodi FS, Fisher DE: **From genes to drugs: targeted strategies for melanoma.** *Nat Publ Group* 2012, **12**:349–361.
2. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, Garnett MJ, Bottomley W, Davis N, Dicks E, Ewing R, Floyd Y, Gray K, Hall S, Hawes R, Hughes J, Kosmidou V, Menzies A, Mould C, Parker A, Stevens C, Watt S, Hooper S, Wilson R, Jayatilake H, Gusterson BA, Cooper C, Shipley J, *et al*: **Mutations of the BRAF gene in human cancer.** *Nature* 2002, **417**:949–954.
3. Pollock PM, Harper UL, Hansen KS, Yudt LM, Stark M, Robbins CM, Moses TY, Hostetter G, Wagner U, Kakareka J, Salem G, Pohida T, Heenan P, Duray P, Kallioniemi O, Hayward NK, Trent JM, Meltzer PS: **High frequency of BRAF mutations in nevi.** *Nat Genet* 2002, **33**:19–20.
4. Pollock PM, Harper UL, Hansen KS, Yudt LM, Stark M, Robbins CM, Moses TY, Hostetter G, Wagner U, Kakareka J, Salem G, Pohida T, Heenan P, Duray P, Kallioniemi O, Hayward NK, Trent JM, Meltzer PS: **High frequency of BRAF mutations in nevi.** *Nat Genet* 2003, **33**:19–20.
5. Patton EE, Widlund HR, Kutok JL, Kopani KR, Amatruda JF, Murphey RD, Berghmans S, Mayhall EA, Traver D, Fletcher CDM: **BRAF Mutations Are Sufficient to Promote Nevus Formation and Cooperate with p53 in the Genesis of Melanoma.** *Curr Biol* 2005, **15**:249–254.
6. Dovey M, White RM, Zon LI: **Oncogenic NRAS Cooperates with p53 Loss to Generate Melanoma in Zebrafish.** *Zebrafish* 2009, **6**:397–404.
7. Huang FW, Hodis E, Xu MJ, Kryukov GV, Chin L, Garraway LA: **Highly Recurrent TERT Promoter Mutations in Human Melanoma.** *Science* 2013, **339**:957–959.
8. Horn S, Figl A, Rachakonda PS, Fischer C, Sucker A, Gast A, Kadel S, Moll I, Nagore E, Hemminki K, Schadendorf D, Kumar R: **TERT Promoter Mutations in Familial and Sporadic Melanoma.** *Science* 2013, **339**:959–961.
9. Krauthammer M, Kong Y, Ha BH, Evans P, Bacchiocchi A, McCusker JP, Cheng E, Davis MJ, Goh G, Choi M, Ariyan S, Narayan D, Dutton-Regester K, Capatana A, Holman EC,

Bosenberg M, Sznol M, Kluger HM, Brash DE, Stern DF, Materin MA, Lo RS, Mane S, Ma S, Kidd KK, Hayward NK, Lifton RP, Schlessinger J, Boggon TJ, Halaban R: **Exome sequencing identifies recurrent somatic RAC1 mutations in melanoma.** *Nat Genet* 2012, **44**:1006–1014.

10. Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, Theurillat J-P, Nickerson E, Auclair D, Li L, Place C, DiCara D, Ramos AH, Lawrence MS, Cibulskis K, Sivachenko A, Voet D, Saksena G, Stransky N, Onofrio RC, Winckler W, Ardlie K, Wagle N, Wargo J, Chong K, Morton DL, Stemke-Hale K, Chen G, Noble M, Meyerson M, Ladbury JE, *et al*: **A landscape of driver mutations in melanoma.** *Cell* 2012, **150**:251–263.

11. Berger MF, Hodis E, Heffernan TP, Deribe YL, Lawrence MS, Protopopov A, Ivanova E, Watson IR, Nickerson E, Ghosh P, Zhang H, Zeid R, Ren X, Cibulskis K, Sivachenko AY, Wagle N, Sucker A, Sougnez C, Onofrio R, Ambrogio L, Auclair D, Fennell T, Carter SL, Drier Y, Stojanov P, Singer MA, Voet D, Jing R, Saksena G, Barretina J, *et al*: **Melanoma genome sequencing reveals frequent PREX2 mutations.** *Nature* 2012, **485**:502–506.

12. Berghmans S, Look TA: **tp53 mutant zebrafish develop malignant peripheral nerve sheath tumors.** *Proc Natl Acad Sci U S A* 2005, **102**:407–412.

13. Sharpless NE, DePinho RA: **The INK4A/ARF locus and its two gene products.** *Curr Opin Genet Dev* 1999, **9**:22–30.

14. Johnson SL, Nguyen AN, Lister JA: **Mitfa is required at multiple stages of melanocyte differentiation but not to establish the melanocyte stem cell.** *Dev Biol* 2011, **350**:405–413.

15. Lister JA, Capper A, Zeng Z, Mathers ME, Richardson J, Paranthaman K, Jackson IJ, Patton EE: **A Conditional Zebrafish MITF Mutation Reveals MITF levels are Critical for Melanoma Promotion Versus Regression in vivo.** *J Invest Dermatol* 2013:1–22.

16. Lister JA, Robertson CP, Lepage T, Johnson SL, Raible DW: **Nacre encodes a zebrafish microphthalmia-related protein that regulates neural-crest-derived pigment cell fate.** *Development* 1999, **126**:3757–3767.

17. Faucherre A, Taylor GS, Overvoorde J, Dixon JE, Hertog JD: **Zebrafish pten genes have overlapping and non-redundant functions in tumorigenesis and embryonic development.** *Oncogene* 2007, **27**:1079–1086.

18. Ceol CJ, Houvras Y, Jane-Valbuena J, Bilodeau S, Orlando DA, Battisti V, Fritsch L, Lin WM, Hollmann TJ, Ferré F, Bourque C, Burke CJ, Turner L, Uong A, Johnson LA, Beroukhim R, Mermel CH, Loda M, Ait-Si-Ali S, Garraway LA, Young RA, Zon LI: **The histone methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset.** *Nature* 2011, **471**:513–517.

19. Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, McLaren S, Sealy I, Caccamo M, Churcher C, Scott C, Barrett JC, Koch R, Rauch G-J, White S, Chow W, Kilian B, Quintais LT, Guerra-Assuncao JA, Zhou Y, Gu Y, Yen J, Vogel J-H, Eyre T, Redmond S, Banerjee R, *et al*: **The zebrafish reference genome sequence and its relationship to the human genome.** *Nature* 2013, **496**:498–503.

20. Varela I, Tarpey P, Raine K, Huang D, Ong CK, Stephens P, Davies H, Jones D, Lin M-L, Teague J, Bignell G, Butler A, Cho J, Dalgliesh GL, Galappaththige D, Greenman C, Hardy C, Jia M, Latimer C, Lau KW, Marshall J, McLaren S, Menzies A, Mudie L, Stebbings L, Largaespada DA, Wessels LFA, Richard S, Kahnoski RJ, Anema J, *et al*: **Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma.** *Nature* 2011, **569**:539–542.
21. Larson DE, Harris CC, Chen K, Koboldt DC, Abbott TE, Dooling DJ, Ley TJ, Mardis ER, Wilson RK, Ding L: **SomaticSniper: identification of somatic point mutations in whole genome sequencing data.** *Bioinformatics* 2012, **28**:311–317.
22. Simpson JT, Durbin R: **Efficient de novo assembly of large genomes using compressed data structures.** *Genome Res* 2012, **22**:549–556.
23. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z: **Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads.** *Bioinformatics* 2009, **25**:2865–2871.
24. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR: **A census of human cancer genes.** *Nat Rev Cancer* 2004, **4**:177–183.
25. Wei X, Walia V, Lin JC, Teer JK, Prickett TD, Gartner J, Davis S, Stemke-Hale K, Davies MA, Gershenwald JE, Robinson W, Robinson S, Rosenberg SA, Samuels Y: **Exome sequencing identifies GRIN2A as frequently mutated in melanoma.** *Nat Genet* 2011, **43**:442–446.
26. Nikolaev SI, Rimoldi D, Iseli C, Valsesia A, Robyr D, Gehrig C, Harshman K, Guipponi M, Bukach O, Zoete V, Michielin O, Muehlethaler K, Speiser D, Beckmann JS, Xenarios I, Halazonetis TD, Jongeneel CV, Stevenson BJ, Antonarakis SE: **Exome sequencing identifies recurrent somatic MAP2K1 and MAP2K2 mutations in melanoma.** *Nat Genet* 2011, **44**:133–139.
27. Stark MS, Woods SL, Gartside MG, Bonazzi VF, Dutton-Regester K, Aoude LG, Chow D, Sereduk C, Niemi NM, Tang N, Ellis JJ, Reid J, Zismann V, Tyagi S, Muzny D, Newsham I, Wu Y, Palmer JM, Pollak T, Youngkin D, Brooks BR, Lanagan C, Schmidt CW, Kobe B, MacKeigan JP, Yin H, Brown KM, Gibbs R, Trent J, Hayward NK: **Frequent somatic mutations in MAP3K5 and MAP3K9 in metastatic melanoma identified by exome sequencing.** *Nat Genet* 2011, **44**:165–169.
28. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, Nik-Zainal S, Martin S, Varela I, Bignell GR, Yates LR, Papaemmanuil E, Beare D, Butler A, Cheverton A, Gamble J, Hinton J, Jia M, Jayakumar A, Jones D, Latimer C, Lau KW, McLaren S, McBride DJ, Menzies A, Mudie L, Raine K, Rad R, Chapman MS, Teague J, *et al*: **The landscape of cancer genes and mutational processes in breast cancer.** *Nature* 2012, **486**:400–404.
29. Pleasance ED, Stephens PJ, O'Meara S, McBride DJ, Meynert A, Jones D, Lin M-L, Beare D, Lau KW, Greenman C, Varela I, Nik-Zainal S, Davies HR, Ordóñez GR, Mudie LJ, Latimer C, Edkins S, Stebbings L, Chen L, Jia M, Leroy C, Marshall J, Menzies A, Butler A, Teague JW, Mangion J, Sun YA, McLaughlin SF, Peckham HE, Tsung EF, *et al*: **A small-**

cell lung cancer genome with complex signatures of tobacco exposure. *Nature* 2010, **463**:184–190.

30. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, Jones D, Hinton J, Marshall J, Stebbings LA, Menzies A, Martin S, Leung K, Chen L, Leroy C, Ramakrishna M, Rance R, Lau KW, Mudie LJ, Varela I, McBride DJ, Bignell GR, Cooke SL, Shlien A, Gamble J, Whitmore I, Maddison M, Tarpey PS, Davies HR, Papaemmanuil E, *et al*: **Mutational Processes Molding the Genomes of 21 Breast Cancers.** *Cell* 2012, **149**:979–993.

31. Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B, Refsland EW, Kotandeniya D, Tretyakova N, Nikas JB, Yee D, Temiz NA, Donohue DE, McDougale RM, Brown WL, Law EK, Harris RS: **APOBEC3B is an enzymatic source of mutation in breast cancer.** *Nature* 2013, **494**:366–370.

32. Dancyger AM, King JJ, Quinlan MJ, Fifield H, Tucker S, Saunders HL, Berru M, Magor BG, Martin A, Larijani M: **Differences in the enzymatic efficiency of human and bony fish AID are mediated by a single residue in the C terminus modulating single-stranded DNA binding.** *FASEB J* 2012, **26**:1517–1525.

33. Zhu Z, He X, Johnson C, Stoops J, Eaker AE, Stoffer DS, Bell A, Zarnegar R, DeFrances MC: **PI3K is negatively regulated by PIK3IP1, a novel p110 interacting protein.** *Biochem Biophys Res Commun* 2007, **358**:66–72.

34. Davies MA: **The role of the PI3K-AKT pathway in melanoma.** *Cancer J* 2012, **18**:142–147.

35. Van Loo P, Nordgard SH, Lingjærde OC, Russnes HG, Rye IH, Sun W, Weigman VJ, Marynen P, Zetterberg A, Naume B, Perou CM, Børresen-Dale A-L, Kristensen VN: **Allele-specific copy number analysis of tumors.** *Proc Natl Acad Sci* 2010, **107**:16910–16915.

36. Garraway LA, Widlund HR, Rubin MA, Getz G, Berger AJ, Ramaswamy S, Beroukhim R, Milner DA, Granter SR, Du J, Lee C, Wagner SN, Li C, Golub TR, Rimm DL, Meyerson ML, Fisher DE, Sellers WR: **Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma.** *Nature* 2005, **436**:117–122.

37. Litman GW, Hawke NA, Yoder JA: **Novel immune-type receptor genes.** *Immunol Rev* 2001, **181**:250–259.

38. Yoder JA, Mueller MG, Wei S, Corliss BC, Prather DM, Willis T, Litman RT, Djeu JY, Litman GW: **Immune-type receptor genes in zebrafish share genetic and functional properties with genes encoded by the mammalian leukocyte receptor cluster.** *Proc Natl Acad Sci* 2001, **98**:6771–6776.

39. Miller AJ, Mihm MCJ: **Melanoma.** *N Engl J Med* 2006, **355**:51–65.

40. Mitra D, Luo X, Morgan A, Wang J, Hoang MP, Lo J, Guerrero CR, Lennerz JK, Mihm MC, Wargo JA, Robinson KC, Devi SP, Vanover JC, D'Orazio JA, McMahon M, Bosenberg MW, Haigis KM, Haber DA, Wang Y, Fisher DE: **An ultraviolet-radiation-independent**

pathway to melanoma carcinogenesis in the red hair/fair skin background. *Nature* 2012, **491**:413–417.

41. Varela I, Klijn C, Stephens PJ, Mudie LJ, Stebbings L, Galappaththige D, van der Gulden H, Schut E, Klarenbeek S, Campbell PJ, Wessels LF, Stratton MR, Jonkers J, Futreal PA, Adams DJ: **Somatic structural rearrangements in genetically engineered mouse mammary tumors.** *Genome Biol* 2010, **11**:R100.

42. Wartman LD, Larson DE, Xiang Z, Ding L, Chen K, Lin L, Cahan P, Klco JM, Welch JS, Li C, Payton JE, Uy GL, Varghese N, Ries RE, Hooch M, Koboldt DC, McLellan MD, Schmidt H, Fulton RS, Abbott RM, Cook L, McGrath SD, Fan X, Dukes AF, Vickery T, Kalicki J, Lamprecht TL, Graubert TA, Tomasson MH, Mardis ER, *et al*: **Sequencing a mouse acute promyelocytic leukemia genome reveals genetic events relevant for disease progression.** *J Clin Invest* 2011, **121**:1445–1455.

43. Taskén K, Solberg R, Zhao Y, Hansson V, Jahnsen T, Siciliano MJ: **The Gene Encoding the Catalytic Subunit α of cAMP-Dependent Protein Kinase (Locus PRKACA) Localizes to Human Chromosome Region 19p13.1.** *Genomics* 1996, **36**:535–538.

44. D'Orazio J, Fisher DE: **Central role for cAMP signaling in pigmentation and UV resistance.** *Cell Cycle* 2011, **10**:8–9.

45. Boikos SA, Stratakis CA: **Molecular genetics of the cAMP-dependent protein kinase pathway and of sporadic pituitary tumorigenesis.** *Hum Mol Genet* 2007, **16**:R80–R87.

46. Kirschner LS: **PRKAR1A and the evolution of pituitary tumors.** *Mol Cell Endocrinol* 2010, **326**:3–7.

47. Kirschner LS, Carney JA, Pack SD, Taymans SE, Giatzakis C, Cho YS, Cho-Chung YS, Stratakis CA: **Mutations of the gene encoding the protein kinase A type I- α regulatory subunit in patients with the Carney complex.** *Nat Genet* 2000, **26**:89–92.

48. Carney JA: **Carney complex: the complex of myxomas, spotty pigmentation, endocrine overactivity, and schwannomas.** *Semin Dermatol* 1995, **14**:90–98.

49. Busca R, Ballotti R: **Cyclic AMP a key messenger in the regulation of skin pigmentation.** *Pigment Cell Res* 2000, **13**:60–69.

50. March HN, Rust AG, Wright NA, Hoeve ten J, de Ridder J, Eldridge M, van der Weyden L, Berns A, Gadiot J, Uren A, Kemp R, Arends MJ, Wessels LFA, Winton DJ, Adams DJ: **Insertional mutagenesis identifies multiple networks of cooperating genes driving intestinal tumorigenesis.** *Nat Genet* 2011, **43**:1202–1209.

51. Heng L: **wgsim - Read simulator for next generation sequencing.** 2012 2012. <https://github.com/lh3/wgsim>.

52. Kinsella RJ, Kahari A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J, Staines D, Derwent P, Kerhornou A, Kersey P, Flicek P: **Ensembl BioMart: a hub for data retrieval across taxonomic space.** *Database (Oxford)* 2011, **2011**:bar030.

53. Kettleborough RNW, Busch-Nentwich EM, Harvey SA, Dooley CM, de Bruijn E, van Eeden F, Sealy I, White RJ, Herd C, Nijman IJ, Fenyes F, Mehroke S, Scahill C, Gibbons R, Wali N, Carruthers S, Hall A, Yen J, Cuppen E, Stemple DL: **A systematic genome-wide analysis of zebrafish protein-coding gene function.** *Nature* 2014, **496**:494–497.
54. Li H, Durbin R: **Fast and accurate long-read alignment with Burrows-Wheeler transform.** *Bioinformatics* 2010, **26**:589–595.
55. **SomaticSniper scripts.** <https://github.com/genome/somatic-sniper/blob/master/src/scripts/>.
56. *Nexus Copy Number - BioDiscovery - Copy Number Variation.* <http://installers.biodiscovery.com/>.
57. Thorvaldsdottir H, Robinson JT, Mesirov JP: **Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration.** *Brief Bioinformatics* 2013, **14**:178–192.
58. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP: **Integrative genomics viewer.** *Nat Biotechnol* 2011, **29**:24–26.
59. Baudis M, Cleary ML: **Progenetix.net: an online repository for molecular cytogenetic aberration data.** *Bioinformatics* 2001, **17**:1228–1229.
60. Roche DNA Microarray Support <http://www.nimblegen.com/products/software/deva/index.html>.
61. NimbleGen R: **DEVA Software User's Guide v1.2.** *NimbleGen* 2012, **1**:1–200.
62. Nilsen G, Liestol K, Van Loo P, Vollan H, Eide M, Rueda O, Chin S, Russell R, Baumbusch L, Caldas C, Dale ALB, Lingjaerde O: **Copy number: Efficient algorithms for single- and multi-track copy number segmentation.** *BMC Genomics* 2012, **13**:591.
63. Greenman C, Wooster R, Futreal PA, Stratton MR, Easton DF: **Statistical analysis of pathogenicity of somatic mutations in cancer.** *Genetics* 2006, **173**:2187–2198.
64. Pond SLK, Frost SDW, Muse SV: **HyPhy: hypothesis testing using phylogenies.** *Bioinformatics* 2005, **21**:676–679.
65. Cheng J, Wedge DC, Pitt JJ, Russnes HG, Vollan HKM, Pandey BP, Nilsen G, Nord S, White KP, Børresen-Dale A-L, Campbell PJ, Kristensen VN, Stratton MR, Lingjærde OC, Moreau Y, Van Loo P: *The landscape of tumor suppressors in primary tumors.*
66. *Index of gene.* <http://ftp.ncbi.nih.gov/gene/>.
67. Lopes CT, Franz M, Kazi F, Donaldson SL, Morris Q, Bader GD: **Cytoscape Web: an interactive web-based network browser.** *Bioinformatics* 2010, **26**:2347–2348.

Additional files

Additional_file_1 as XLSX

Additional file 1 Table S1 Zebrafish tumours used in the exome study. **Table S2** Sequencing coverage and metrics. **Table S3** Somatic mutations identified in the 53 zebrafish melanomas. **Table S4** Copy number changes identified in the 53 zebrafish melanomas. **Table S5** P-values of genes occurring in amplifications. **Table S6** P-values of genes occurring in homozygous deletions. **Table S7** Insertion and deletions identified in the 53 zebrafish melanomas. **Table S8** Statistical Analysis of Mutation Burden Correlation. **Table S9** Significance of genes with respect to frequency and modality. **Table S10** Mutated pathways and their significance from enrichment analysis.

Additional_file_2 as PDF

Additional file 2 Figure S1: Effect of SNP density on germline and somatic substitution calling performance using CaVEMan. **Figure S2:** Comparison of substitution calling algorithms on zebrafish melanoma data. **Figure S3:** Experimental outline. **Figure S4:** Evidence of two additional cluster of mutations in ZD8a on chromosome 10. **Figure S5:** Comparison of copy number aberration profiles between ASCAT and aCGH. **Figure S6:** Unsupervised clustering analysis of copy number aberrations. **Figure S7:** Pathway analysis of all mutations. **Figure S8:** Distribution of co-occurring copy number alterations and/or somatic mutations in TP53, MITF, and CDKN2A across 120 BRAF mutant melanomas identified in the SKCM TCGA dataset.

Additional_file_3 as DOCX

Additional file 3 Supplementary text describing the mutation calling simulations and comparison of mutation callers.

BRAF^{V600E};p53^{-/-}



NRAS^{Q61K};p53^{-/-}

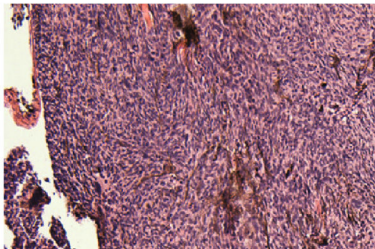
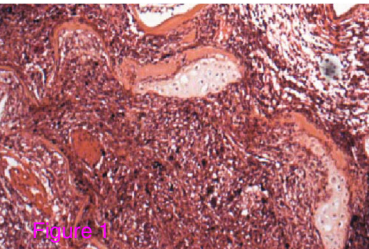
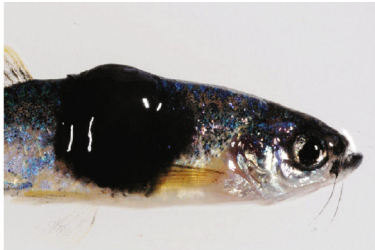


Figure 1

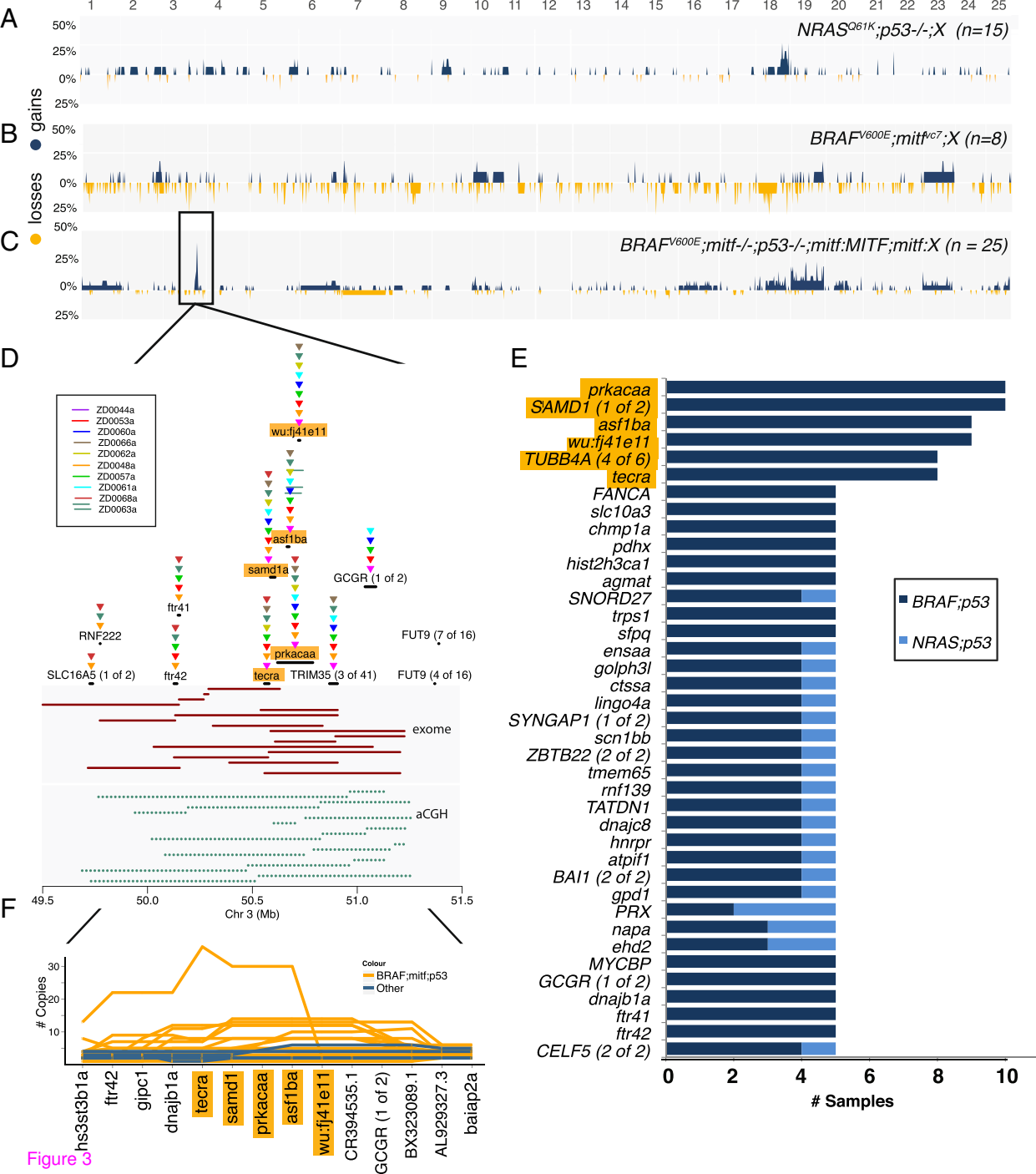
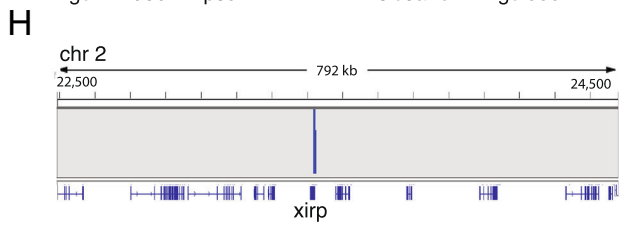
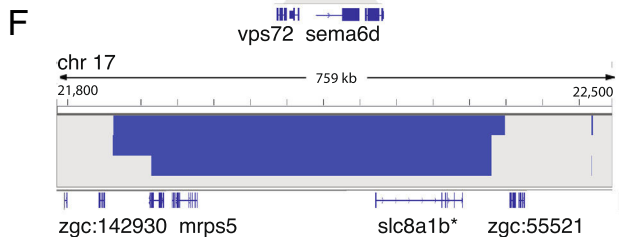
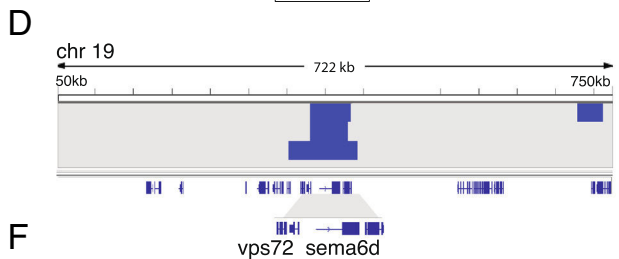
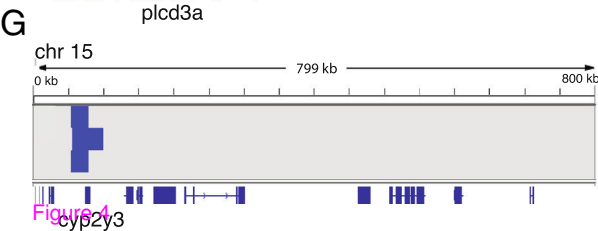
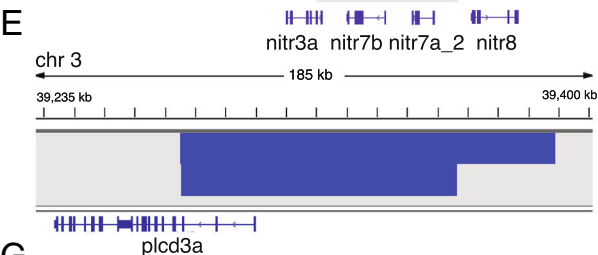
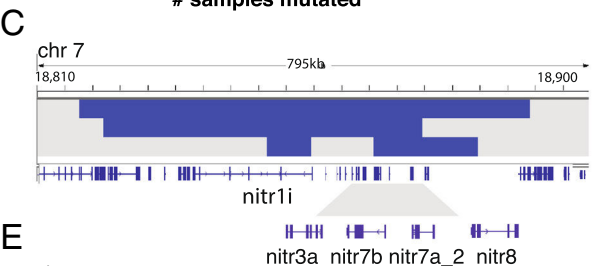
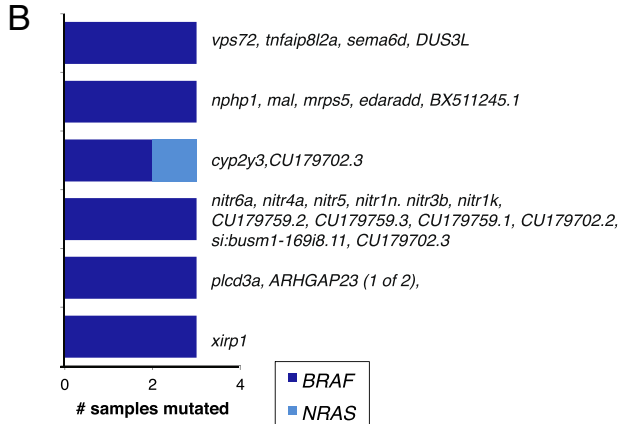
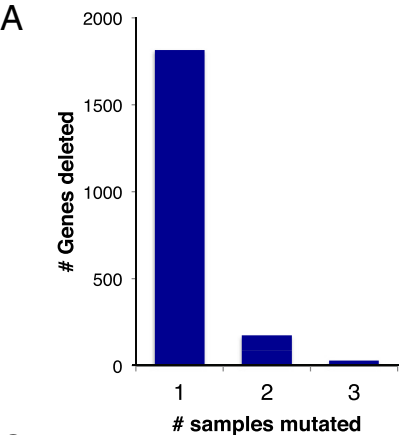


Figure 3



Additional files provided with this submission:

Additional file 1: 1959096090999633_add1.xlsx, 2024K

<http://genomebiology.com/imedia/3560338411110889/supp1.xlsx>

Additional file 2: 1959096090999633_add2.pdf, 3060K

<http://genomebiology.com/imedia/3504430001110889/supp2.pdf>

Additional file 3: 1959096090999633_add3.docx, 134K

<http://genomebiology.com/imedia/3104313381110889/supp3.docx>